VISIT…

LANZAROTE

Caliente.COM

Нейрокомпьютеры: разработка, применение. 2003. №12. - С.???

Робастная целевая функция с допуском на точность решения для нейросети-предиктора

Царегородцев В.Г.
www.NeuroPro.ru tsar@neuropro.ru

Предложено два варианта введения допуска по точности решения задачи в робастную целевую функцию на основе обобщенной степенной метрики. Новая целевая функция применима как для традиционной постановки задачи обучения нейросети-предиктора, так и для задач автоассоциации и clearning-технологий обучения.

Robust cost function based on Minkowski power metric is extended by introducing allowable approximation errors for predicting variables. Area of usage includes classical regression task performed by multilayer perceptron and tasks of autoassociation and clearning.

Целевые функции для задачи нейропрогнозирования

Для обучения нейросети алгоритмом обратного распространения необходимо иметь возможность оценивать близость выходного сигнала, выдаваемого сетью в ответ на некоторое входное воздействие, к требуемому. Критерии вторичной оптимизации, например, регуляризующие, в работе не рассматриваются.

С момента появления алгоритма обратного распространения в 1986г. в качестве целевой функции наиболее широко используется МНК-функция $f(y,\tilde{y}) = \frac{1}{2}(y-\tilde{y})^2$, где y и $\tilde{y}$ – требуемое и выданное на выходе сети скалярные значения соответственно. Множитель 1/2 традиционно возникает для простоты выражения первых производных. В случае векторного выхода нейросети берется сумма по всем выходным сигналам: $F(\mathbf{y},\tilde{\mathbf{y}}) = \sum_i f(y_i,\tilde{y}_i)$. Значение целевой функции для набора обучающих примеров вычисляется суммированием по примерам выборки: $G = \sum_j F(\mathbf{y}^j,\tilde{\mathbf{y}}^j)$. Возможно также введение устанавливаемых пользователем весов для каждого примера выборки [1].

В обзорных работах по алгоритмам обучения нейросетей наблюдается разнобой: одни могут подчеркивать важность адекватного выбора целевой функции, например, [2], другие – нет [3]. В монографической литературе наблюдается та же картина.

Использование МНК-оценки не является оптимальным для многих случаев. Для задачи классификации с учителем специальные целевые функции были разработаны преимущественно для случая наличия сигмоидных нелинейных функций, биполярных или униполярных, у нейронов выходного слоя сети – это кросс-энтропийные функции [4,5], энтропийная функция [6], функции [7,8] для предотвращения насыщения сигнала выходного нейрона при правильном ответе и ускорения выхода из области насыщения при неправильном ответе сети. Ссылки на другие работы есть в [9-11]. Сам же автор на практике использует функции типа расстояния до множества правильно интерпретируемых ответов [1,12].

Для задач аппроксимации и прогнозирования (регрессия, авторегрессия, автоассоциация) наиболее отличным от МНК является каскадно-корреляционный подход [13] понейронного наращивания нейросети, когда при обучении синапсов добавляемого в скрытый слой нейрона максимизируется модуль корреляции его выходного сигнала и ошибки нейронов выходного слоя (после завершения обучения нейрона веса синапсов нейронов выходного слоя обучаются с использованием МНК-

оценки). Манипулирование структурой сети происходит и в задаче нелинейной автоассоциации [14] для максимальной компрессии данных каждой из нелинейных главных компонент, но выходные сигналы тоже оцениваются МНК- функцией.

Другие же варианты целевых функций для задачи прогноза, фактически, образуют две группы – робастные целевые функции и функции с возможностью задания допуска на точность решения задачи. В этих рамках и будет идти изложение.

Робастные функции для обучения нейропредиктора

Для задач нейроаппроксимации наиболее ранней из предложенных отличных от МНК функций была робастная целевая функция [15] на основе обобщенной степенной метрики Минковского:

$$f(y,\tilde{y}) = \frac{1}{r}\,|\,y-\tilde{y}\,|^{r}\ . \qquad (1)$$

Эта робастная функция взята из разработанного в классической статистике подхода [16] и, по сравнению с МНК, при $1 \le r < 2$ является более адекватной для распределений переменных с тяжелыми хвостами, т.е. в негауссовском случае. Увеличение же степени метрики, т.е. r>2, при отсутствии шума в данных может ускорить обучение на начальных итерациях, как замечено в [17].

Иной статистический подход по обеспечению робастности к выбросам в данных был предложен [18] и успешно опробован с нейросетями в [19]. Здесь также использовано снижение степени метрики, но проведено разделение на "малые" и "большие" отклонения модели от данных: малые отклонения по прежнему штрафуются МНК-оценкой, для больших используется l_1-метрика:

$$f(y,\tilde{y}) = \begin{cases} \dfrac{1}{2}(y-\tilde{y})^2, & \text{\it если} \quad |\,y-\tilde{y}\,| \le \varepsilon, \\ \varepsilon\,|\,y-\tilde{y}\,| - \dfrac{\varepsilon^2}{2}, & \text{\it если} \quad |\,y-\tilde{y}\,| > \varepsilon. \end{cases}$$

Обе эти робастные функции можно напрямую использовать при обучении сети. Особо нужно отметить, что они удовлетворяют теоретическим требованиям (например, [20]) на вид целевой функции, самостоятельно не порождающей дополнительных ложных локальных минимумов: отсутствуют разрывы как самой функции, так и ее производной. Многие же предложенные в нейроинформатике целевые функции (например, [11,21]) как раз грешат такими разрывностями.

Для достижения робастности используются не только специальные целевые функции, но и предварительное исправление выборок данных. Так, в [22] для коррекции выбросов в данных выборки кластеризуются, и примеры, статистически достоверно выходящие за типичный радиус разброса своего кластера, проецируются по направлению к центру кластера.

Ввод допуска в робастную функцию на основе метрики Минковского

Если зависимые переменные имеют высокую грубость инструментальных измерений, можно прекращать обучение сети при достижении величин отклонений, по модулю меньших или равных половине инструментальной единицы измерения. Для этого можно использовать целевую функцию, имеющую нулевое значение при допустимых отклонениях и штрафующую отклонения за пределами допуска.

Для нейросетей такая модификация для МНК-функции была предложена [21,23], однако, имела разрыв как самой функции, так и ее производной:

$$f(y,\tilde{y})=\begin{cases} 0. & если \quad |y-\tilde{y}|\le\varepsilon, \\ \frac{1}{2}(y-\tilde{y})^2, & если \quad |y-\tilde{y}|>\varepsilon. \end{cases}$$

Здесь ε – константа, задающая допуск по точности для *y*. МНК-функция с допуском и отсутствием разрывов дана в [12].

Введем возможность задания допуска на точность аппроксимации значений переменной в робастную целевую функцию (1) на базе степенной метрики:

$$f(y,\tilde{y})=\begin{cases} 0, & если \quad |y-\tilde{y}|\le\varepsilon, \\ \frac{1}{r}(|y-\tilde{y}|-\varepsilon)^r, & если \quad |y-\tilde{y}|>\varepsilon. \end{cases} \quad (2)$$

$$f(y,\tilde{y})=\begin{cases} 0, & если \quad |y-\tilde{y}|\le\varepsilon, \\ \frac{1}{r}\left(\frac{|y-\tilde{y}|}{\varepsilon}-1\right)^r, & если \quad |y-\tilde{y}|>\varepsilon. \end{cases} \quad (3)$$

Вариант (2) можно назвать "абсолютным" вариантом, вариант (3) – "относительным". Для задач с единственной зависимой переменной могут применяться оба варианта. Формула (3) может быть лучшей при предсказании векторного значения (векторная предикция, автоассоциация), поскольку выражает отклонение каждой переменной в долях от требуемого для нее допуска: при сильно различающихся допусках для различных зависимых переменных можно уравнять вклады невязок для этих переменных в суммарную целевую функцию примера и далее в обучение нейросети (естественно, сравнивать надо величины допусков, полученные для предобработанных зависимых переменных – т.е. в случае линейного масштабирования диапазонов значений зависимых признаков в приемлемые для нейросети диапазоны допуски тоже нужно перемасштабировать умножением на отношения длин исходных и полученных после предобработки интервалов значений переменных).

Практическое применение

Результаты нейромоделирования с использованием робастных целевых функций на основе степенной метрики и хьюберовского подхода представлены в [15,19]. Дополнительно можно отметить следующие моменты.

- Целевая функция с допуском позволяет автоматически прекращать процесс обучения нейросети при достижении нужной точности описания данных. При этом остается возможным параллельное использование внешних критериев ранней остановки обучения, например, по достижению минимума ошибки обобщения.
- В задаче автоассоциации при значительных инструментальных погрешностях измерений целевая функция с допуском снижает требования к точности описания данных нейросетью и может приводить к получению меньшего числа линейных или нелинейных главных компонент (сигналов в среднем, "узком" слое нейросети).
- Можно вести обучение, т.е. вычисление значений оптимизируемой функции и градиента, при меньших допусках, останавливая обучение при стабильном решении всех примеров с реально нужной точностью на протяжении нескольких последних итераций. При этом обученная сеть будет находиться в области flat minima [24], где небольшие возмущения адаптивных параметров сети не будут выводить точность решения за исходный допуск, даже без применения специальных алгоритмов обучения наподобие предложенного в [24].
- В [12] предложен способ поиска примеров-"диссидентов" и примеров с выбросами в данных повторением этапов исключения примера с наибольшей ошибкой и дообучения сети. Снижение степени метрики в целевой функции увеличивает

робастность этой эмпирики: исключенные разными нейросетями наборы примеров порождают меньшее число примеров в объединении и бóльшее число примеров в пересечении этих наборов, чем при случае использования МНК-оценки (с допуском или без него).

- Подход *clearning* [25] одновременного обучения сети и коррекции входных данных использует для примера целевую функцию вида $H = F(\mathbf{y}, \tilde{\mathbf{y}}) + \lambda F(\mathbf{x}, \tilde{\mathbf{x}})$, где второе слагаемое штрафует отклонение реальных значений независимых переменных от "желаемых" их значений, по мнению модели. Значения $\tilde{\mathbf{x}}$ добавляются к вектору параметров модели и корректируются с учетом вычисленного алгоритмом обратного распространения градиента для примера. Эта идея отличается от традиционного подхода первоначальной коррекции данных и последующего построения модели (пример – вышеописанный метод [22]). Робастную целевую функцию с допуском можно использовать и здесь, как для зависимых, так и для независимых переменных, получая при снижении степени метрики (по сравнению с МНК-случаем) более надежное исправление отдельных сильных выбросов в независимых переменных.

Литература

1. *Миркес Е.М.* Нейрокомпьютер: проект стандарта. Новосибирск: Наука, 1999. - 337с.
2. *Frasconi P., Gori M., Tesi A.* Successes and failures of backpropagation: a theoretical investigation / Progress in Neural Networks. Ablex Publishing, 1993, Vol.5. - pp.205-242.
3. *LeCun Y., Bottou L., Orr G.B., Müller K.-R.* Efficient BackProp / Neural Networks: Tricks of the trade (G.Orr and K.Müller, eds.), Springer Lecture Notes in Comp. Sci. 1524, 1998. - pp.5-50.
4. *Solla S.A., Levin E., Fleisher M.* Accelerated learning in layered neural networks / Complex Systems, 1988, Vol.2. - pp.625-639.
5. *Hertz J., Krogh A., Palmer R.G.* Introduction to the theory of neural computation. Addison Wesley, 1991.
6. *Kamimura R.* Unification of information maximization and minimization / Advances in Neural Information Processing Systems 9 (1996). MIT Press, 1997.– pp.508-514.
7. *Oh S.-H., Lee Y.* A modified error function to improve the error back-propagation algorithm for multi-layer perceptrons / ETRI Journal, Vol.17, №1, April 1995. - pp.11-21.
8. *Lee H.-M., Huang T.-C., Chen C.-M.* Learning efficiency improvement of back propagation algorithm by error saturation prevention method / Proc. IJCNN'1999, Washington, DC,USA. - 6p.
9. *Lister R., Bakker P., Wiles J.* Error signals, exceptions and back propagation / Proc. IJCNN'1993, Nagoya, Japan, 1993. Vol.1. - pp.573-576.
10. *Dumitras A., Munteanu C., Lazaresku V.* Nover cost functions for neural classifiers / Proc. Eur. Congress on Intelligent Techniques and Soft Computing (EUFIT'97), Aachen, Germany, 1997. Vol.1. - pp.475-479.
11. *Falas T., Stafylopatis A.-G.* The impact of error function selection in neural network-based classifiers / Proc. IJCNN'1999, Washington, DC, USA. - 6p.
12. *Горбань А.Н., Россиев Д.А.* Нейронные сети на персональном компьютере. Новосибирск: Наука, 1996. - 276с.
13. *Fahlmann S.E., Lebiere C.* The cascade-correlation learning architecture / Advances in Neural Information Processing Systems 2 (1989). Morgan Kaufmann, 1990. - pp.524-532.
14. *Takahashi T., Tokunaga R.* Energy functions for efficient nonlinear dimensionality reduction by multi layer perceptrons / Proc. Int. Conf. Neural Information Processing, 1998, Vol.1. - pp.494-497.
15. *Hanson S.J., Burr D.J.* Minkowski-r back-propagation: learning in connectionist models with non-euclidean error signals / Advances in Neural Information Processing Systems 0 (1987). American Institute of Physics, 1988. - pp.348-357.
16. *Mosteller F., Tukey J.* Robust estimation procedures. Addison Wesley, 1980.
17. *Rohwer R.* The 'moving targets' training algorithm / Advances in Neural Information Processing Systems 2 (1989). Morgan Kaufmann, 1990. - pp.558-565.
18. *Huber P.J.* Robust statistics. Wiley, 1981. (Русский перевод: *Хьюбер П.* Робастность в

статистике. - М.: Мир, 1984).
19. *Xi C., Wang F., Devabhaktuni V.K., Zhang Q.-J.* Huber optimization of neural networks: a robust training method / Proc. IJCNN'1999, Washington, DC, USA. - 4p.
20. *Hamey L.G.C.* Result on weight configuration that are not local minima in feed-forward neural networks / Proc. VII Australian Conf. Artificial Neural Networks, 1996. - pp.173-178.
21. *Bakker P.* Don't care margins help back-propagation learn exceptions / Proc. Australian Conf. Artificial Intelligence, World Scientific, 1992. - pp.139-144.
22. *Hämäläinen J.J., Järvimäki I.* Input projection method for safe use of neural networks based on process data / Proc. IJCNN'1998, Anchorage, Alaska, USA, 1998. - pp.193-198.
23. *Watanabe E., Nakayama H.* A learning algorithm for avoidance of overfittng in function approximation by neural networks / Proc. ICNN'1994, Seoul, Korea, 1994. Vol.1. - pp.219-224.
24. *Hochreiter S., Schmidhuber J.* Simplifying neural nets by discovering flat minima / Advances in Neural Information Processing Systems 7 (1994). MIT Press, 1995. – pp.529-536.
25. *Weigend A.S., Zimmermann H.G., Neuneier R.* The observer-observation dilemma in neuro-forecasting: reliable models from unreliable data through CLEARNING / Artificial Intelligence on Wall Street (R.Freedman, ed.), Software Engineering Press, NY, USA, 1995. - pp.308-317.